Multivariate abundance model

Julie Vercelloni

Model description

A multivariate abundance model is developed to estimate the influences of environmental predictors including management zoning and Degree Heating Weeks (DHW) on changes in abundance of benthic groups across Karimumjawa National Park, Indonesia.

We adopted the joint species modelling approach to tease apart effects of observed environmental predictors on correlated multivariate responses (Warton et al., 2015; Hui, 2016). Correlations are induced via latent variables to capture the residual variability not accounted by model predictors. Latent variables are unknown and estimated by the model which allow to reduce the number of model parameters to be estimated in a multivariate framework (Warton et al., 2015). Latent variables are seen as unmeasured environmental predictors, or as ordination scores, capturing the co-variation of community abundance in a low-dimensional space after controlling for the observed predictors.

This approach creates new opportunities to predict abundance and co-occurrence patterns across many biological communities and understand drivers of changes at broad spatial scales. Unlike many multivariate distance-based analyses, the joint modelling approach accounts for the statistical properties of the data which can bring serious problems if ignored (Warton et al., 2012).

The abundance of benthic groups was characterized as relative counts for each subtransect i, benthic group j and time t. Counts were modelled using a negative binomial distribution (**NB**) parameterized with a mean parameter (μ_{ij}) and an over-dispersion parameter (ϕ_j) specific to the benthic group. These parameters were linked to environmental and latent variables via a log-link (Eq. 1).

$$
\begin{aligned}
y_{ij} &\sim \mathbf{NB}(\mu_{ij}, \phi_j) \\
log(\mu_{ij}) &= \alpha_i + \beta_{0_j} + \theta_i + x_i^T \beta_j + z_i^T \lambda_j
\end{aligned}
\tag{1}
$$

with β_j and λ_j are vectors of coefficients for each benthic group related to the environmental and latent variables, respectively.

The environmental component, $x_i^T \beta_j$, is composed of three environmental predictors and an intercept β_{j_0} that accounts for differences in functional group abundance (Eq. 2).

$$
x_i^T \beta_j = \beta_{j_0} + \beta_{j_1} Zoning_i + \beta_{j_2} DHW_i + \beta_{j_3} Year_i
\tag{2}
$$

The latent component, $z_i^T \lambda_j$, is composed of two correlated latent variables z_i^T formulated as random parameters and loading factors λ_j (Eq. 3). In addition, subtransect-level intercepts (α_i) are modelled as a random effect ($\alpha_i \sim \mathbf{N}(0, \sigma^2)$). The parameter allows to include the compositional aspect of the data by standardizing all terms in the model for the total number of counts (Hui et al., 2015) that summed to 50 as per the photo-quadrat processing (González-Rivero et al., 2016). This parameter also include the repeated measurements of subtransect between years.

$$
z_i^T \lambda_j = z_{i_1} \lambda_{j_1} + z_{i_2} \lambda_{j_2}
\tag{3}
$$

The Bayesian aspect of the model obliges to indicate prior distributions on model parameters. Co-occurrences were parameterized using multivariate normal distributions with mean equal to 0 and covariance $c_0\mathbf{I}$ (Eq. 4).

$$
\begin{aligned}
z_{i_1}, z_{i_2} &\sim \mathscr{N}(0, \mathbf{I}) \\
\beta_{j_0}, \beta_{j_1}, \beta_{j_2}, \beta_{j_3}, \alpha_i, \lambda_{j_1}, \lambda_{j_2} &\sim \mathscr{N}(0, c_0\mathbf{I}) \\
\phi &\sim \mathscr{U}(0, 10) \\
c_0 &= 10
\end{aligned}
\tag{4}
$$

Co-occurrence patterns of benthic groups were estimated from the residual correlation from the covariance matrix. We interpreted these correlations as evidence of co-occurrence related to interactions not explained by our observed environmental predictors. Interactions due to shared responses to environmental predictors were estimated by calculating the covariance between model estimates of environmental predictors. Co-occurrence is expressed in terms of significant correlations varying between -1 and 1 where 95% credible intervals of their posterior distributions did not include zero. The R package Boral was used to implement the multivariate abundance model (Hui, 2016).

The multivariate abundance model is fit on two different ecological grouping. The first grouping includes five benthic categories (Hard coral, Soft coral, Algae, Invertebrates and Other) as presented in the main manuscript. The second grouping divides the communities into 15 sub-groups based on morphological traits (Table S1).

Model selection

The best model formulation was selected from five models using the Watanabe–Akaike information criterion (wAIC). Model formulation associated with the minimum value of wAIC is kept. (Table S1).

Table S1: Ranked model selection based on Conditional DIC values using the 5 benthic groups as response variables.

Model	Features	WAIC	Conditional DIC
Model 4	Correlated responses, environmental predictors, repeated measurements	30841.1	32624.2
Model 3	Correlated responses, environmental predictors	31276.6	31974.3
Model 1	Correlated responses	32646.5	32965.3
Model 2	Independent responses, environmental predictors	35768.4	32758.8

Table S2: Ranked model selection based on WAIC values using the 15 sub-groups as response variables.

Model	Features	WAIC	Conditional DIC
Model 4	Correlated responses, environmental predictors, repeated measurements	63917.9	64698.9
Model 3	Correlated responses, environmental predictors	64395.1	66478.4
Model 1	Correlated responses	67901.1	65162.6
Model 2	Independent responses, environmental predictors	70026.7	69105.5

Model validation

Model validation was assessed by examining Bayesian posterior predictive distributions of the benthic groups (and sub-groups). The discrepancy distributions were estimated from the differences between posterior predictive distributions and observed counts for each MCMC simulation (Figs. S1 and S2). These distributions were used to compute posterior predictive p-values and root mean squared errors (RMSE, Tables S3 and S4). These assessments can be compared to cross-correlation approaches typically used for model validation with the added benefit of being able to be implemented directly from model outputs. Model goodness-of-fit was also diagnosed by plotting observed values against discrepancy distributions, overall predicted versus observed data and model residual distributions per functional group.

Five benthic groups

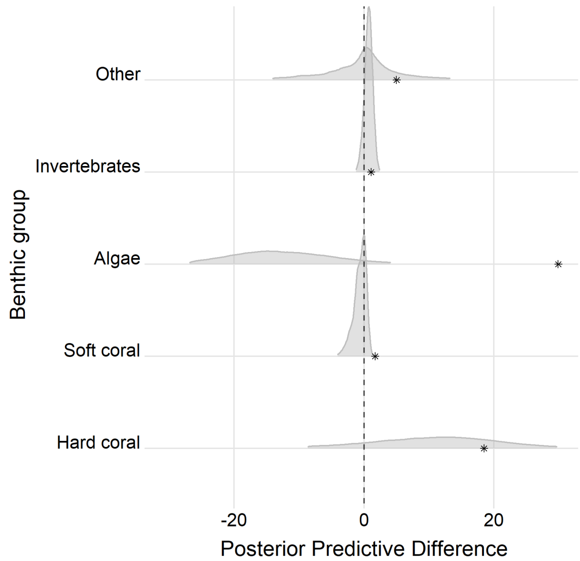
Table S3: Values of model-goodnesses-of-fit diagnostics.

Benthic group	RMSE	p-value
Hard coral	13.7	0.862
Soft coral	1.07	0.297
Algae	13.9	0.118
Invertebrates	0.808	0.881
Other	6.20	0.552

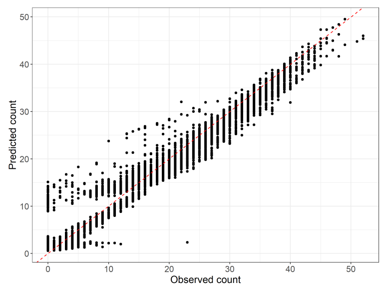
Fifteen benthic groups

Table S4: Values of model-goodnesses-of-fit diagnostics.

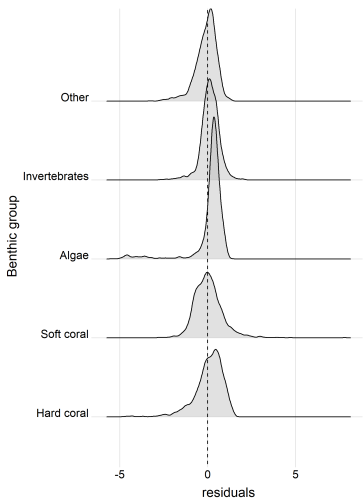
Benthic sub-group	RMSE	p-value
ACR_BRA	29.9	0.544
ACR_TCD	29.3	0.591
BRA_nACR	30.9	0.450
CCA	29.5	0.851
EAM	71.0	0.033
FLP	30.5	0.223
FREE	28.4	0.636
GORG	29.0	0.889
MACRO	29.1	0.673
MSE	31.5	0.183
MSEM	28.4	0.641
NON_COLONIZABLE	34.7	0.294
NON_HERM	28.6	0.721
OTH-SF	28.1	0.481
SPONG	28.1	0.557



(a) Discrepancy distributions

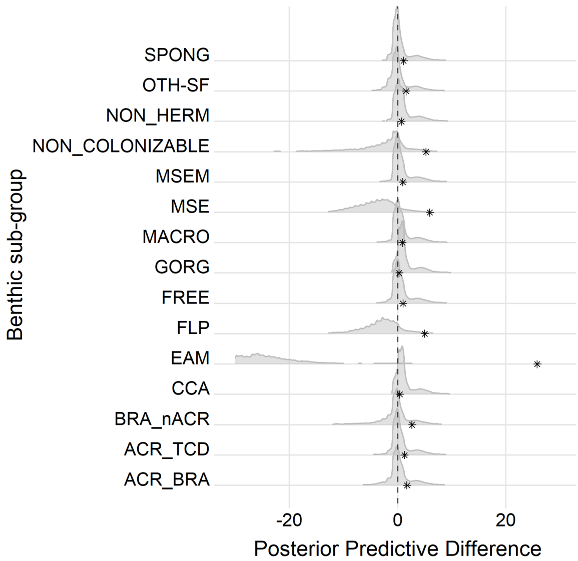


(b) Overall model fit

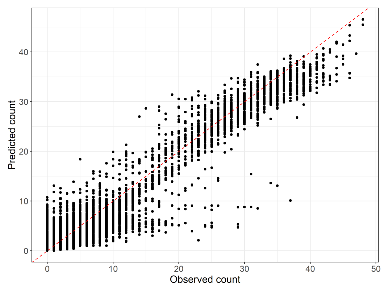


(c) Residual distributions

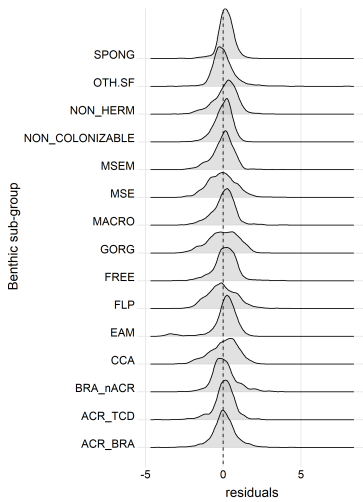
Figure S1: Model performance of benthic groups. (a) the distributions show differences between predictions and observed data from MCMC simulations. Asterisks indicate the average counts derived from the data.



(a) Discrepancy distributions



(b) Overall model fit



(c) Residual distributions

Figure S2: Model performance of benthic sub-groups. (a) the distributions show differences between predictions and observed data from MCMC simulations. Asterisks indicate the average counts derived from the data.

References

González-Rivero, M., O. Beijbom, A. Rodriguez-Ramirez, T. Holtrop, Y. González-Marrero, A. Ganase, C. Roelfsema, S. Phinn, and O. Hoegh-Guldberg. 2016. Scaling up ecological measurements of coral reefs using semi-automated field image collection and analysis. Remote Sensing .

Hui, F. K. 2016. boral–Bayesian ordination and regression analysis of multivariate abundance data in R. Methods in Ecology and Evolution **7**:744–750.

Hui, F. K., S. Taskinen, S. Pledger, S. D. Foster, and D. I. Warton. 2015. Model-based approaches to unconstrained ordination. Methods in Ecology and Evolution **6**:399–411.

Warton, D. I., F. G. Blanchet, R. B. O'Hara, O. Ovaskainen, S. Taskinen, S. C. Walker, and F. K. Hui. 2015. So many variables: joint modeling in community ecology. Trends in Ecology & Evolution **30**:766–779.

Warton, D. I., S. T. Wright, and Y. Wang. 2012. Distance-based multivariate analyses confound location and dispersion effects. Methods in Ecology and Evolution **3**:89–101.